Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Smith TJ, Kahaly GJ, Ezra DG, et al. Teprotumumab for thyroid-associated ophthalmopathy. N Engl J Med 2017;376:1748-61. DOI: 10.1056/NEJMoa1614949

Teprotumumab for Thyroid-Associated Ophthalmopathy

Terry J. Smith, et al.

Supplemental Materials

Table of Contents

Item	Content		Page
1	Text	Details of scales used to assess efficacy	2
2	Table S1	Schedule of assessments	3
3	Table S2	Additional patient characteristics	5
4	Table S3	Randomization by clinical center	6
5	Table S4	Changes in $HgbA1_c$	7
6	Table S5	Chi square CMH test – responder analysis	9
7	References	Four citations from the Text (Item 1)	9

(1) Details of Scales Used to Assess Efficacy

Clinical Activity Score

The clinical activity score consists of seven components: spontaneous retrobulbar pain, pain on attempted eye movements (upward, side-to-side, and downward gazes), conjunctival redness, redness of the eyelids, chemosis, swelling of the caruncle/plica, and swelling of the eyelids.1 Each component is scored as present or absent, 1 or 0. The score at each efficacy assessment is the sum of all items present; i.e. giving a range of 0-7, where 0 or 1 constitutes inactive disease and 7 severe active ophthalmopathy. A change of $\geq$2 points is considered clinically meaningful.

Graves' Ophthalmopathy Quality of Life (GO-QoL)

Quality of life was evaluated with the use of the Graves' ophthalmopathy quality of life questionnaire.[2,3] The questionnaire has two self-assessment subscales; one covering impact of visual function on daily activities, the other assesses the impact of self-perceived appearance. The visual function subscale covers activities such as driving, walking outdoors, reading, watching television. The appearance subscale asks questions such as whether ophthalmopathy has altered your appearance, caused people to have a negative reaction to you, caused social isolation, and caused you to try to mask your appearance. Each subscale has 8 questions which are answered either; yes – very much so, yes – a little, or no – not at all. Each question is scored 0-2 and the total raw score is then mathematically transformed to a 0-100 scale, where 0 represents the most negative impact on quality of life, and 100 represents no impact. A change of $\geq$8 points on the 0-100 scales has been shown to be clinically meaningful. The combined score takes raw scores from both subscales and again transforms them to a single 0-100 scale.

Gorman Grading of Diplopia

The Gorman assessment of subjective diplopia includes four categories: no diplopia (absent), diplopia when the patient is tired or awakening (intermittent), diplopia at extremes of gaze (inconstant), and continuous diplopia in the primary or reading position (constant).[4] Patients are scored according to which grade of diplopia they are experiencing. An improvement of $\geq$1 grade is considered clinically meaningful.

(2) Table S1. Schedule of Assessments

	Screening[1]		Treatment Period											EW
	→													
Study Visit	SV1→3		1[2]	2	3[2]	4	5	6	7	8	9	10	11	
Week (W)	W−4		W0	W1	W3	W4	W6	W9	W12	W15	W18	W21	W24	EW1
Measure	±2 W		±3 d	± 1 d	± 3 d	± 1 d	± 3 d	± 3 d	± 3 d	± 3 d	± 3 d	± 3 d	± 3 d	
Physical & ophthalmic exam[3]	X		X	X	X	X	X	X	X	X	X	X	X	X
ECG[4]	X		X		X		X		X				X	X
Lab tests[5]	X		X	X[6]	X	X[6]	X	X	X	X[6]	X	X[6]	X	X
Infusion			X		X		X	X	X	X	X	X		
CAS[7]	X		X				X		X		X		X	X
Subjective diplopia[7]	X		X				X		X		X		X	X
Proptosis[7]	X		X				X		X		X		X	X
Biomarkers[8]			X						X				X	X
PK samples[9]			X	X	X	X		X					X	
AE[10]			X	X	X	X	X	X	X	X	X	X	X	X
GO-QOL	X						X		X				X	X
ADA			X		X			X					X	

Randomization

Abbreviations: ADA, anti-drug antibodies; AE, adverse event; CAS, clinical activity score; d, day; ECG, electrocardiogram; EW, early withdrawal; GO-QoL, Graves' ophthalmopathy quality of life questionnaire; PK, pharmacokinetics; SV, screening visit; W, week.

[1] Written informed consent obtained at, or prior to, the screening visit; prior to any study-related procedures being performed. Medical history, including smoking history and patient demographics, was collected at screening visit.

[2] Research staff contacted patients the day after the first and second infusions to inquire about safety and tolerability and thereafter only if any infusion-related AE had occurred previously.

[3] Ophthalmic exam included best corrected visual acuity, pupil exam, color vision assessment, Ishihara color plates (or equivalent) or related red desaturation, intraocular pressure and slit lamp exam. If significant abnormalities were noted as compared to previous visits, further investigations of visual function were conducted according to the ophthalmologist's judgment.

[4] ECGs were performed in triplicate, one minute apart at all time points. At weeks 0, 6 and 12 only, ECGs were performed prior to and following the infusion.

[5] Hematology, biochemistry, FT3, FT4, urinalysis and pregnancy test (serum at screening then urine thereafter) prior to dose at each time point except weeks 1, 4, 15, and 21, when only hematologic parameters and glucose were measured. Hemoglobin A_{1c} was performed at screening, weeks 12 and 24.

[6] Hematology and blood glucose assessments only at weeks 1, 4, 15, and 21 for all patients. Patients were instructed to fast prior to blood collection at weeks 1 and 4.

[7] Performed prior to dose.

[8] Samples collected for the inflammatory cytokines interleukin-6, interleukin-16, and Regulated on Activation, Normal T Cell Expressed (RANTES).

[9] Sampling performed prior to infusion and at the end of infusion at weeks 0, 3, and 9. A single sample drawn at weeks 1, 4 and 24.

[10] Treatment emergent AEs were reported following the start of the first infusion.

[11] Graves' ophthalmopathy quality of life (GO-Qol) questionnaires completed by patients.

[12] Serum anti-drug antibodies (ADA) were assessed prior to dosing at weeks 0, 3, 9 and 24.

(3) Table S2. Additional Characteristics of Patients

Patient Characteristics[1]		
Demographic	**Placebo (n=44)**	**Teprotumumab (n=43)**
Race – number of patients (%)		
White	38 (86.4)	37 (86.0)
Asian	2 (4.5)	1 (2.3)
Black	4 (9.1)	4 (9.3)
Previous thyroid treatment – number of patients (%)		
Radioiodine	5 (11.4)	5 (11.6)
Surgical thyroidectomy	5 (11.3)	5 (11.6)
Number of patients with prior steroid therapy (%)	1 (2.3)	0
$HgbA_{1c}$ (n)[1]	5.6 ± 0.5 (37)	5.6 ± 0.5 (36)
Number of patients with diabetes at baseline (%)	6 (13.6)	6 (14.0)
Liver function tests; outside normal range at baseline[2] (%)		
ALT	5 (11)	5 (12)
AST	5 (11)	3 (7)
ALP	12 (27)	12 (28)
Renal function tests; outside normal range at baseline[3] (%)		
Blood urea nitrogen / Creatinine	2 (4.5)	2 (4.7)
Study eye – number of patients (%)[8]		
Left eye	24 (54.5)	16 (37.2)
Right eye	20 (45.4)	27 (62.8)

Abbreviations: ALT, alanine amino transferase; AST, aspartate amino transferase; ALP, alkaline phosphatase; $HgbA_{1c}$, glycosylated hemoglobin A_{1c}; ITT, intent to treat population.

[1] Numbers were lower than ITT because baseline $HgbA_{1c}$ was not captured for some of the early patients; assay was added and procedures were changed through a protocol amendment.

[2] Number patients outside normal range at baseline. For ALT and AST all patients were ≤ 3 times upper limit of normal (an inclusion criteria), 2 placebo patients were > 2 times upper limit of normal. For ALP two placebo patients were > 2 times and one patient was > 3 times the upper limit of normal.

[3] No clinically significant deviations outside normal range.

(4) Table S3. Randomization by Clinical Center

US Clinical Center Code	Number Patients Randomized	Percentage Patients Randomized	EU Clinical Center Code	Number Patients Randomized	Percentage Patients Randomized
003	11	12.5	**050**	19	21.6
004	10	11.4	**054**	10	11.4
029	10	11.4	**051**	3	3.4
001	8	9.1	**054**	3	3.4
022	4	3.4			
016	3	3.4			
016	2	2.3			
023	2	2.3			
015	1	1.1			
020	1	1.1			
021	1	1.1			
Total	**53**	**60.2**		**35**	**39.8**

For calculating percentages, a total of 88 patients were randomized. One patient withdrew consent prior to dosing and was not included in the ITT population.

(5) Table S4. Changes in $HgbA_{1c}$

Site#	Patient #	Screening	Week 12	Week 24	Week 36	Week 72
Placebo						
001	**0001**	NA	NA	NA	6.2	6.4
029	**0001**	NA	NA	NA	6.2	5.5
003	**0001**	NA	NA	NA	8.0	6.9
053	**0003***	6.9	6.6	7.6	8.0	6.9
050	**0001***	6.3	6.5	6.0	6.0	6.3
050	**0003**	6.2	6.1	5.9	5.8	5.8
050	**0008**	6.2	6.1	6.6	6.6	7.1
Teprotumumab						
021	**0001***	6.2[#]	NA	NA	NA	6.2
023	**0001***	6.1[#]	NA	NA	6.6	6.3
004	**0003***	7.2[#]	NA	7.0	7.6	6.4
015	**0001***	7.2	7.3	Early termination (6 infusions)		
022	**0002***	6.4	6.2	6.3	6.1	6.5
003	**0009***	7.4	8.2	7.8	7.3	7.7
050	**0010**	6.3	5.8	5.9	6.2	5.8
013	**0002**	5.9	6.3	6.4	6.3	6.0

Table captures all trial patients with multiple elevations in blood glucose having $HgbA_{1c}$ levels that were between 6.1-6.4% (elevated risk for diabetes), or $\geq$ 6.5% (diabetic). If multiple measurements were made at any time point, the highest value is shown in the table. *, patients who were diagnosed as well-controlled diabetics on study entry; [#], values obtained from medical records prior to screening; NA, value not available ($HgbA_{1c}$ assays were started after the trial was initiated through a protocol amendment, thus some values for early patients are missing).

Summary of diabetes medications and adjustments; no information for a patient indicates no medications.

Placebo: 003-0001, metformin 500 mb BID started week 32; 053-0003, study entry metformin 500 mg QD, increased to 500 mg BID week 36, 30 mg glicazide added week 60; 050-0001, study entry metformin 1000 mg BID, 200 mg sitagliptin QD added week 4; 050-0003, study entry metformin 500 mg BID; 050-0008, study entry metformin 100 mg BID and sitaglipin 50 mg BID.

Teprotumumab: 004-0003, study entry metformin 500 mg QD, 5 mg glipizide added week 6, 023-0001, study entry metformin 500 mg and liraglutide 0.6 mg QD, liraglutide increased to 1.2 mg for 3 weeks at week 15, dulagutide 2mg/week added at week 18; 015-0001, diagnosed as diabetic on screening, study

entry metformin 500 mg BID, increased to 1000 mg week 3, glipizide 2.5 mg BID added week 15; 022-002, study entry insulin glargine 50 U QPM and insulin aspart 10-10-14 U TID, no changes; 003-0009, diagnosed diabetic on screening, study entry metformin 500 mg QD, increased to 500 mg BID week 3 and 1000 mg BID week 18, glimepiride 2 mg QD added week 18; 050-0010, started metformin 500 mg BID week 6, stopped after 2 weeks.

(6) Table S5: Chi-square / Cochran–Mantel–Haenszel Responder Analysis

Parameter / Analysis	Placebo (n = 45)	Teprotumumab (n = 42)
Responder	9 (23.1)[1]	29 (76.3)
Non-responder	30 (76.9)	9 (23.7)
Total	39 (100)	38 (100)
Missing	6	4
Chi-square Test[2]		
Difference		53.2%
95% CI		34.3%, 72.1%
P-value		<0.001
CMH Test[3]		
Odds ratio (teprotumumab vs. placebo)		10.37
95% CI		3.57, 30.15
P-value		<0.001

[1], values in parentheses are % of total (missing week 24 excluded).
[2], Chi-square test comparing week 24 responder vs. non-responder; Difference = responder proportion in teprotumumab – responder proportion in placebo.
[3], Cochran–Mantel–Haenszel test stratified on baseline smoking status (non-smoker vs. smoker)

(7) References

1. Mourits MP, Prummel MF, Wiersinga WM, Koornneef L. Clinical activity score as a guide in the management of patients with Graves' ophthalmopathy. Clin Endocrinol (Oxf) 1997;47:9-14. [Erratum, Clin Endocrinol (Oxf) 1997;47:632.]

2. Terwee CB, Gerding MN, Dekker FW, Prummel MF, Wiersinga WM. Development of a disease specific quality of life questionnaire for patients with Graves' ophthalmopathy: the GO-QOL. Br J Ophthalmol 1998;82:773-9.

3. Terwee CB, Dekker FW, Mourits MP, et al. Interpretation and validity of changes in scores on the Graves' Ophthalmopathy Quality of Life Questionnaire (GO-QOL) after different treatments. Clin Endocrinol (Oxf) 2001;54:391-8.

4. Bahn RS, Gorman CA. Choice of therapy and criteria for assessing treatment outcome in thyroid-associated ophthalmopathy. Endocrinol Metab Clin North Am 1987;16:391-407.